Fine-Grained Detection of Inverse Tone Mapping in HDR Images

Wei Fan, Giuseppe Valenzise, Francesco Banterle, Frédéric Dufaux

Fine-Grained Detection of Inverse Tone Mapping in HDR Images

Wei Fan[a], Giuseppe Valenzise[b], Francesco Banterle[c], Frédéric Dufaux[b]

[a] *Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA*
[b] *Laboratoire des Signaux et Systèmes (L2S, UMR 8506), CNRS - CentraleSupélec - Université Paris-Sud, 91192 Gif-sur-Yvette, France*
[c] *Visual Computing Lab, ISTI-CNR, Pisa, Italy*

Abstract

High dynamic range (HDR) imaging enables to capture the full range of physical luminance of a real-world scene, and is expected to progressively replace traditional low dynamic range (LDR) pictures and videos. Despite the increasing HDR popularity, very little attention has been devoted to new forensic problems that are characteristic to this content. In this paper, we address for the first time such kind of problem, by identifying the source of an HDR picture. Specifically, we consider the two currently most common techniques to generate an HDR image: by fusing multiple LDR images with different exposure time, or by inverse tone mapping an LDR picture. We show that, in order to apply conventional forensic tools to HDR images, they need to be properly preprocessed, and we propose and evaluate a few simple HDR forensic preprocessing strategies for this purpose. In addition, we propose a new forensic feature based on Fisher scores, calculated under Gaussian mixture models. We show that the proposed feature outperforms the popular SPAM features in classifying the HDR image source on image blocks as small as 3×3 pixels, which makes our method suitable to detect composite forgeries combining HDR patches originating from different acquisition processes.

☆The work presented in this document was supported by BPIFrance and Région Ile de France, in the framework of the FUI 18 Plein Phare project. Part of the work was done when W. Fan, G. Valenzise, and F. Dufaux were with LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay.

Keywords: Digital image forensics, high dynamic range imaging, inverse tone mapping, Fisher scores

1. Introduction

One key limitation of conventional 8-bit image representations is that the range of luminance available in the physical world substantially exceeds the dynamic range handled by traditional imaging pipelines, resulting in a loss of visual information in over/under-exposed image regions. This bottleneck was initially recognized and studied in computer graphics [1, 2, 3], leading to the development of high dynamic range (HDR) formats capable of storing the physical luminance of a scene (expressed in cd/m^2) in high bit-depth, floating-point formats. More recently, HDR image and video formats have been popularized in multimedia applications [4], thanks to the increasing availability of HDR cameras and displays, as well as content compression standards [5, 6]. Current trends in multimedia technology, such as the diffusion of Ultra High Definition Television [7], seem to suggest that the HDR paradigm will progressively replace conventional low dynamic range (LDR) imaging [8].

While one can reasonably expect that native HDR sensors will become available at low cost in the next coming years, nowadays the two most common techniques to generate HDR content include: 1) acquiring multiple conventional LDR pictures of the scene at different exposure times, which can be fused together afterwards using, e.g., the method in [1]. We will refer to HDR pictures generated in this way as **mHDR**; 2) acquiring an LDR picture of the scene, and expanding its dynamic range through an operation commonly known as *inverse tone mapping* (iTM), since conceptually it does the opposite of tone mapping algorithms conceived to display HDR pictures on LDR displays [9]. We will refer to this kind of images as **iHDR**. This latter option is particularly attractive considering that nowadays the majority of legacy video footage is LDR, and that range expansion is needed to display it on next-generation HDR displays [10]. Furthermore, it has been shown that in many cases HDR video obtained through

iTM yields similar, or even indistinguishable, visual experience as HDR content generated by multiple exposures [11, 12].

In this work, we consider the forensic problem of identifying whether an HDR picture is mHDR or iHDR. From a multimedia security perspective, solving this forensic problem can help to identify the authenticity of a content, and to localize tampering whenever mHDR and iHDR image patches have been composed together to create a forgery. To this end, it is desirable to perform fine-grained mHDR/iHDR classification to precisely localize the tampered regions and infer their semantics consequently [13].

Forensic problems in the context of HDR images have been rarely considered in the literature. In part, this is due to the only recent development of HDR imaging in the multimedia and signal processing community; at the same time, the limited availability of HDR image datasets has somehow constrained forensic research in this field; finally, the very same concept of high dynamic range image format has sometimes been erroneously confused with simply higher bit depth (sometimes called wider dynamic range), whereas HDR pictures are scene referred and represent real-world luminance, featuring very different statistical characteristics compared to LDR images. As a result, only very limited work has been done to identify and solve new potential forensic problems associated to HDR, and when this has been done, it was in the LDR domain, *e.g.*, to differentiate LDR images from tone-mapped HDR images [14]. To the best of our knowledge, this is the first work targeting a forensic problem *in the HDR domain*, *i.e.*, analyzing directly HDR pictures in order to identify their source.

Indeed, in spite of the similarities between HDR and LDR imagery, HDR image forensics present some subtleties and new challenges with respect to standard forensic techniques. For instance, while iTM might resemble a contrast enhancement process, classical forensic detectors based on statistical fingerprints [15, 16, 17] fail when applied on iHDR pictures, as those images do not present typical peak/gap artifacts in their histogram. In fact, the histogram of the HDR image is not composed by a fixed number of bins, *e.g.*, 2^8; conversely, the bin size can be arbitrarily chosen, and the maximum luminance value is

content dependent. As a result of the continuous nature of HDR values, most forensic methods based on integer arithmetics, such as classical compression and point-wise transformation detectors [18, 19, 16] are not applicable. More sophisticated forensic tools, which employ higher-order statistics to model local image content [20, 21], are based on comparing neighboring pixel values as if they were lying on a uniform interval scale. While this is a reasonable assumption in the case of LDR pixels, we show in this paper that HDR images need to be preprocessed through a non-linear transformation in order for these forensic tools to provide acceptable results. This confirms previous findings in HDR image quality assessment [22, 23] and in local feature extraction from HDR pictures [24, 25].

Though as an extension of our preliminary work [26], this paper departs from [26] on the following four aspects. 1) A comprehensive study is carried out for HDR image preprocessing, covering different strategies and parameters. 2) Instead of 8-bit representation, we consider a much more challenging case when iHDR images are obtained from high bit-depth RAW LDR images with 16 bits integer precision. 3) We provide an in-depth study of inverse tone mapping detection on very small image blocks, including the smallest possible size 3×3. 4) The number of selected inverse tone mapping algorithms has increased from 3 to 6. In all, as the first work addressing HDR image forensics, we have made the following contributions to differentiate mHDR and iHDR images:

- As a starting point, we consider the peculiarities of HDR content compared to LDR images, from a forensic perspective. We introduce and motivate some basic preprocessing steps that must be applied to HDR content in order to extract meaningful forensic features. Differently from preprocessing approaches proposed in other fields of HDR imaging such as quality evaluation and compression [22, 27], the preprocessing techniques considered in this work are specifically designed and evaluated to maximize forensic classification accuracy rather than perceptual fidelity.

- Based on the analysis of joint histograms of mHDR and iHDR images,

we observe that features describing higher-order statistics of local image patches are necessary to differentiate mHDR from iHDR. Based on this observation, we propose an HDR image forensic feature based on the Fisher scores [28, 29] calculated under parametric Gaussian mixture models (GMMs). Our approach is to some extent inspired by a texture/facial analysis feature [30, 31], which however has never been applied in a forensic scenario. We show that our feature can outperform the popular Subtractive Pixel Adjacency Matrix (SPAM) features [20] in mHDR and iHDR image classification.

- In order to address the problem of fine-grained mHDR/iHDR classification for, *e.g.*, tampering localization, we conduct an experimental study using image blocks with *very small* sizes, including the challenging case of 3×3 image blocks. This shows that the proposed forensic algorithm has a practical value in detecting composition forgeries including iHDR and mHDR content.

The rest of the paper is organized as follows. Sec. 2 briefly reviews the mHDR image creation and six iTM methods. In Sec. 3, we compare LDR and HDR image representations from a forensic point of view, and propose four simple strategies to pre-process an HDR image in order to effectively extract forensic features. The proposed method based on Fisher scores and GMMs is presented in Sec. 4. Finally, conclusions are drawn in Sec. 6.

2. Background

In this paper, we consider a new image forensic problem on HDR content: differentiating an HDR image created from the fusion of multiple LDR images (*i.e.*, mHDR) from one created through inverse tone mapping of a single LDR image (*i.e.*, iHDR). In this section we briefly describe how these two types of images are created.

2.1. mHDR image creation

The most popular way to create HDR images is by combining pictures taken from a conventional LDR camera [1, 32], using multiple LDR photos of the same scene captured with different exposure times. For each captured image, pixel values are modeled as a non-linear camera response function of the exposure. The exposure is further modeled as the multiplication of the scene luminance (to be estimated), and the exposure time (assumed to be known). Therefore, the camera response function and the luminance of the scene can be jointly estimated, *e.g.*, by minimizing the squared error between the camera response and the pixel values measured at the different exposure times. The estimated luminance map constitutes the resulting mHDR image.

2.2. iHDR image creation

The dynamic range of an LDR picture can be expanded to match that of an HDR display by iTM operators. Compared to mHDR images, iHDR images created by iTM operators do not represent the physical luminance of the scene, but rather aim at expanding the dynamics of the LDR image in such a way to reproduce the visceral response associated to the original scene [33]. Subjective studies have shown that, with a well-exposed LDR image and a proper iTM operator, iHDR images can be visually as appealing as mHDR pictures, and thus difficult to distinguish from the latter ones at naked eye [12, 11]. This motivates the study of forensic techniques to computationally assess whether an HDR image is iHDR or mHDR.

In this paper, we select six popular inverse tone mapping algorithms for creating the iHDR images, described in the following.

1) In *Akyüz et al.'s method* [11], denoted by '**A**', the input LDR content is firstly linearized and then linearly scaled to achieve the desired dynamic range (typically the one of the HDR displaying device). Akyüz *et al.* show that this method works well with well-exposed content without compression artifacts.

2) *Banterle et al.'s method* [34, 35], denoted by '**B**', expands content by applying the inverse of a sigmoid [36]. To reduce artifacts due to quantization

and to try to reconstruct the lost signal, an expanded image is blended with the original one using an expand map, *i.e.*, a smooth field that guides expansion based on importance sampling. The method works well when images are moderately saturated [9] and is temporally stable [35].

3) *Huo et al.'s method* [37], denoted by '**H**', uses a non-linear sigmoid-like function in order to increase the dynamic range of the input LDR content only in specific regions of the image, indicated by an expand map. The method is computationally fast (using an approximated bilateral filter), temporally stable, and is able to achieve quality results on moderately over-exposed content.

4) *Kovaleski and Oliveira's method* [38], denoted by '**K**', linearly expands the input LDR content to a desired dynamic range only in certain regions of the image using an expand map, which is computed using thresholding on the luminance channel followed by bilateral filtering. As in Huo *et al.*'s operator [37], the method is computationally fast, temporally stable, and it can achieve quality results on moderately over-exposed content.

5) *Meylan et al.'s method* [39], denoted by '**M**', applies different linear expansions in different areas of the image, which are classified as diffuse or specular. Classification is based on simple thresholding. Filtering is employed to reduce contour artifacts that may appear in areas between a diffuse and a specular area. The method is straightforward to implement and fast, but it is not temporally coherent because thresholding depends on the processed image.

6) *Rempel et al.'s method* [33], denoted by '**R**', linearly expands the input LDR content to a desired dynamic range only in certain regions of the image using an expand map, which is obtained using an edge-stop function computed in a multi-resolution fashion. This method is computationally fast and temporally stable. The method can typically produce high quality results on moderately over-exposed content.

3. Image Forensics: from LDR to HDR

The existing abundant literature on image forensics mainly deals with conventional, 8-bit LDR representation per pixel per color channel. A natural question arises about whether these forensic techniques might be applied to HDR content, and which kinds of precautions and caveats should be considered while doing so. In this section, we start by discussing some fundamental differences between HDR and LDR images, which are often misunderstood or confused. Afterwards, we consider the illustrative case of a popular steganalytic/forensic feature, SPAM [20], and study its applicability in the case of HDR image.

3.1. Differences between HDR and LDR Image Representations

A common interpretation of HDR images is that those are the higher bit-depth version of LDR pictures, *e.g.*, being 16 or more bits per pixel and per channel. However, the difference with LDR is conceptually deeper.

LDR images are *device-referred* or output-referred, *i.e.*, pixel values represent color to be displayed on a monitor or paper print. Since perceived luminance is a nonlinear function of physical light, LDR pixels are perceptually encoded using the sRGB non-linearity [40], commonly approximated as a "gamma-correction" function (with the typical gamma value of 2.2). This gamma correction, which was historically introduced to compensate for the typical response of legacy CRT displays, actually describes quite accurately the loss of sensitivity of the human visual system at low luminance levels. Due to this nonlinear compression, it is possible to represent images to be reproduced on typical LDR displays using 8 bits per pixel per color channel. 16-bit representations are also becoming popular, *e.g.*, as the RAW output of many digital cameras. Furthermore, pixel values in LDR images are approximately perceptually uniform, *i.e.*, a difference of 1 pixel value has approximately the same perceptual magnitude independently of the baseline value on which the difference is computed. As a result, arithmetic operations on LDR pixels are perceptually meaningful. Moreover, due to this perceptual uniformity, a well-contrasted LDR picture has pixel values that span the full range of available codewords (*e.g.,* $0 - 255$).

On the other hand, HDR images are *scene-referred*, *i.e.*, pixel values store (or are proportional to) the physical luminance of the scene, measured in cd/m^2. As mentioned above, real-world luminance may span several orders of magnitude, from as low as 10^{-6} cd/m^2 for very dim scenes, up to 10^8 cd/m^2 for direct sunlight in a sunny day. In order to accurately represent this wide range, HDR image formats such as OpenEXR [41] or RGBE [2] typically use floating point pixel values. Hence, differently from the LDR case, HDR pixels do not take values on a fixed interval, *i.e.*, the minimum and maximum pixel values are arbitrary and can be very different across images. Furthermore, the typical histogram of an HDR image is highly skewed, with most pixels concentrated in a very small range, but with possibly very long tails due to bright details at high luminance. Finally, due to the highly nonlinear human perception of brightness, in order to perform perceptually meaningful arithmetic operations on HDR pixel values, these need to be previously encoded in a similar way as in the case of the gamma correction [22]. Different encoding functions have been proposed in the past few years, including the popular SMPTE 2084 electro-optical transfer function [27].

3.2. Benchmark Feature: SPAM

One of the contributions of this paper is to investigate how a typical forensic feature performs for a novel problem in the context of HDR images. To this end, we consider the well-known steganalytic/forensic feature SPAM (Subtractive Pixel Adjacency Matrix) [20]. In LDR image forensics literature, it has shown excellent performance in detecting various image operations such as JPEG compression [42], median filtering [43], and image sharpening [13]. We choose to compare with the SPAM feature instead of the well-known SRM (34671-dimensional) feature [21] because of the following two reasons. Firstly, the SPAM feature can be taken as a subset of the SRM feature [21]. Secondly, we are interested in forensic detection of image blocks as small as 3×3, whose high-dimensional SRM feature may contain lots of redundancy.

The SPAM feature was designed for 8-bit grayscale images. It models the

image derivatives using the first-order and second-order Markov chains. In this paper, we consider the widely used second-order SPAM feature, which is 686-dimensional. Clearly, as mentioned in Section 3, computing differences between photometric pixel values might not be appropriate, *e.g.*, these differences might be dominated by very bright pixels. Hence, it is reasonable to apply some compressive encoding (such as a logarithm) before computing the forensic feature. In Sec. 5, we will provide relevant experimental comparisons between the SPAM feature and the proposed method.

3.3. HDR Image Preprocessing

Straightforwardly, it is not possible to directly extract the SPAM feature from an HDR image given its floating point representation and largely varying pixel value range. In fact, this can bias the extraction of the 686-dimensional second-order SPAM feature as it only counts small image derivatives with values in $\{-3, -2, \cdots, 3\}$. For this problem, HDR image pixel values should be scaled and rounded so that the SPAM feature can be properly extracted. In the following, we consider four simple scaling and rounding strategies, which will then be compared experimentally in Section 5.2.

Without loss of generality, we only consider the luminance component of a given HDR image, which is also where most iTM methods carry out the dynamic range expansion. Given an HDR image of size $H \times W$, its luminance component is extracted and linearly scaled to $[0, 1]$ to obtain the matrix $\mathbf{L}$. A simple scaling strategy can be based on the maximum pixel value, which is equal to 1 for $\mathbf{L}$. However, this does not take into account that different HDR images may capture scenes with very different dynamic ranges. In order to compensate for this factor, we can scale the image based on the average brightness. To this end, we adopt the image key which indicates whether the captured scene is subjectively light, normal, or dark [36]. This can be approximated as the exponential of the log-average luminance of the image:

$$K = \exp\left(\frac{1}{H \times W} \sum_{i,j} \log\left(\mathbf{L}_{i,j} + \epsilon\right)\right), \tag{1}$$

where $\mathbf{L}_{i,j}$ denotes the value of the (i,j)-th pixel $(1 \leq i \leq H, 1 \leq j \leq W)$, and we set $\epsilon = 10^{-6}$ to avoid the singularity when taking the logarithm. Notice that Eq. (1) corresponds to computing the geometric mean of the image, which is intuitively appropriate given the highly skewed distribution of HDR pixels. Based on either the maximum value 1 or the image key K calculated adaptively according to each image, we thereafter propose the following two image scaling strategies:

$$\mathbf{L}_{i,j}^{lin} = (2^b - 1)\mathbf{L}_{i,j}, \tag{2}$$

and

$$\mathbf{L}_{i,j}^{keylin} = (2^b - 1)\frac{\mathbf{L}_{i,j}}{K}, \tag{3}$$

where b is an arbitrary non-negative parameter. The value of b decides the range of pixel values after scaling, which affects the population of image derivatives in $\{-3, -2, \cdots, 3\}$ considered by the SPAM feature [20]. We will experimentally study its impact on the forensic performance in Sec. 5.2.

In the study of HDR image statistics, it is a common practice to analyze the image after taking the logarithm [44]. Accordingly, we propose the two following strategies to scale the HDR image in the logarithmic domain:

$$\mathbf{L}_{i,j}^{log} = (2^b - 1)\frac{\log(\mathbf{L}_{i,j} + \epsilon) - \log(\epsilon)}{\log(1 + \epsilon) - \log(\epsilon)}, \tag{4}$$

and

$$\mathbf{L}_{i,j}^{keylog} = (2^b - 1)\frac{\log(\mathbf{L}_{i,j} + \epsilon) - \log(\epsilon)}{\log(K + \epsilon) - \log(\epsilon)}, \tag{5}$$

respectively based on the maximum pixel value and the image key. After the HDR image is preprocessed following either Eq. (2), (3), (4), or (5), the pixel values are rounded to integers in order to extract the SPAM feature.

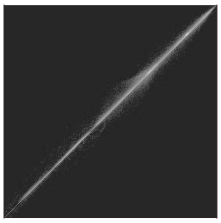
In order to extract the proposed forensic feature (to be detailed in the next section), we also consider the four preprocessing strategies as for SPAM features. However, our method does not require integer pixels. Thus, no rounding is applied after Eqs. (2)-(5) for the extraction of our feature.

4. Forensic Analysis Based on Fisher Scores

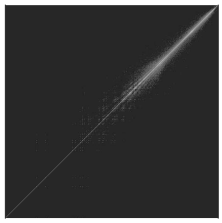
4.1. Preliminary Analysis

In order to discriminate iHDR and mHDR images, a key observation is that iHDR pictures are obtained from an LDR signal, which has a discrete nature. For 8-bit representation LDR images with pixel values in a limited integer set $\{0, 1, \cdots, 255\}$, though after inverse tone mapping the pixel values are converted to floating point numbers, the discreteness of pixel values still exist in the iHDR images. In order to explore these statistics, we analyze the joint probability or co-occurrence matrix of neighboring pixels [20, 45]. When constructing the joint histogram with a sufficiently big number of bins, *e.g.*, 512, such intrinsic differences between mHDR and iHDR images can be exposed.

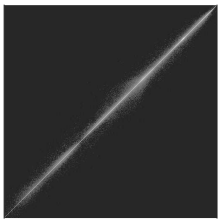
Fig. 1 reports the joint histogram of horizontally adjacent pixels of the luminance component in some HDR images (refer to the electronic version for a better visibility). Differently from the mHDR images, we can observe that gaps/peaks exist in the joint histograms of iHDR images, as shown in Fig. 1-(b), -(d), -(f), -(h), -(j), and -(l). To compactly describe the above mentioned joint histogram characteristic, we can use, *e.g.*, Fourier analysis. In our preliminary tests, this provided us around 85% of iHDR image detection accuracy, indicating that such second-order statistics may be only partially effective [26]. In addition, many digital cameras can also output 16-bit RAW pictures, that could be used as input for an iTM algorithm. In this case, the differences in second-order statistics of pixel values might be too small to be accurately detected, as shown in Fig. 1-(c), -(e), -(g), -(i), -(k), and -(m), where we can no longer observe gaps/peaks in the joint histogram. This suggests that second-order statistics might be insufficient to solve, with acceptable generality, the forensic problem of discriminating iHDR from mHDR. Therefore, we resort to a more powerful feature, based on higher-order statistics, which provides a richer description of inter-pixel dependencies.



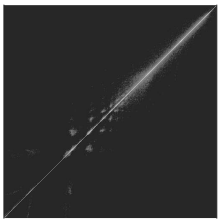
(a) mHDR



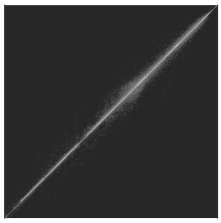
(b) iHDR, '**A**', 8-bit



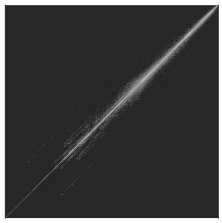
(c) iHDR, '**A**', 16-bit



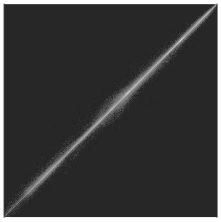
(d) iHDR, '**B**', 8-bit



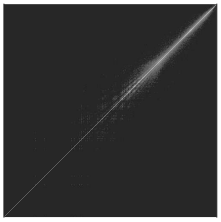
(e) iHDR, '**B**', 16-bit



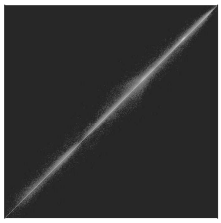
(f) iHDR, '**H**', 8-bit



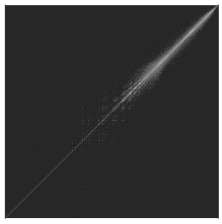
(g) iHDR, '**H**', 16-bit



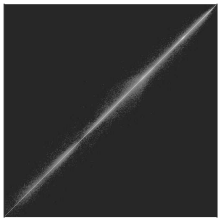
(h) iHDR, '**K**', 8-bit



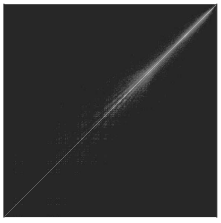
(i) iHDR, '**K**', 16-bit



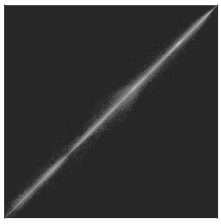
(j) iHDR, '**M**', 8-bit



(k) iHDR, '**M**', 16-bit



(l) iHDR, '**R**', 8-bit



(m) iHDR, '**R**', 16-bit

Figure 1: Example joint histograms of mHDR/iHDR images in the logarithmic domain. The same scene was captured in the mHDR image and in the LDR image which was used for creating the iHDR images [11, 39, 33, 35, 37, 38]. The corresponding LDR image for creating the iHDR images whose joint histograms are shown in (b), (d), (f), (h), (j), and (l) is represented with 8 bits for each color channel. Whereas the corresponding LDR image for (c), (e), (g), (i), (k), and (m) is represented with 16 bits for each color channel. For a better visibility, we have taken logarithm of the joint histogram and afterwards carried out a normalization.

4.2. Forensic Analysis Based on Fisher Scores

Motivated by the preliminary analysis of mHDR/iHDR images in Sec. 4.1, here we study the local higher-order image statistics after the given HDR image is pre-processed as described in Sec. 3.3. Given a generic pixel $\mathbf{z}_0$ and its $s \times s$ local neighborhood $\{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_{s^2-1}\}$, we obtain a local differential vector $\mathbf{x}$ with $\mathbf{x}_i = \mathbf{z}_i - \mathbf{z}_0$ $(i = 1, 2, \cdots, s^2 - 1)$. Its log-likelihood under an M-component Gaussian mixture model (GMM) parametrized by $\boldsymbol{\theta} = \{\pi^i, \boldsymbol{\mu}^i, (\mathbf{C}^i)^{-1} | i = 1, 2, \cdots, M\}$ is computed as:

$$L(\boldsymbol{\theta}|\mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^{M} \pi^i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}^i, \mathbf{C}^i), \tag{6}$$

where π^i, $\boldsymbol{\mu}^i$, and $\mathbf{C}^i$ are respectively the mixing weight, mean, and covariance matrix of the i-th GMM component. The higher-order statistics in the local neighborhood of $\mathbf{z}_0$ can therefore be represented using the Fisher scores [28], which are calculated as the partial derivatives with respect to the parameters $\boldsymbol{\theta}$ of the log-likelihood $L(\boldsymbol{\theta}|\mathbf{x})$, *i.e.*, $\nabla_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}|\mathbf{x})$. More specifically, the partial derivatives with respect to π^i and $(\mathbf{C}^i)^{-1}$ are calculated as:

$$\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{x})}{\partial \pi^i} = \frac{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}^i, \mathbf{C}^i)}{L(\boldsymbol{\theta}|\mathbf{x})}, \tag{7}$$

$$\frac{\partial \log L(\boldsymbol{\theta}|\mathbf{x})}{\partial (\mathbf{C}^i)^{-1}} = \frac{\pi^i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}^i, \mathbf{C}^i)}{2L(\boldsymbol{\theta}|\mathbf{x})} \left(\mathbf{C}^i - (\mathbf{x} - \boldsymbol{\mu}^i)(\mathbf{x} - \boldsymbol{\mu}^i)^T\right). \tag{8}$$

For the sake of simplicity and also for reducing the dimensionality of the final forensic feature, we only consider the diagonal elements of $(\mathbf{C}^i)^{-1}$ for computing the Fisher scores. Therefore, we have a Fisher score vector $\mathbf{F}(\boldsymbol{\theta}, \mathbf{x})$, with length $M + (s^2 - 1)M = s^2 M$.

In practice, we compute the Fisher scores with respect to two 0-mean GMMs, parametrized by $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^1$, representing mHDR and iHDR, respectively. The two GMMs are learned in a previous, off-line training stage from a database containing the two classes of HDR images, using the Expectation-Maximization algorithm. The reason why we employ 0-mean GMMs is that $\mathbf{x}$ captures the local derivatives of $\mathbf{z}$, which carry high-frequency information, and thus the

trained means $\boldsymbol{\mu}^i$ are generally close to zero. Therefore, we remove the DC component and train 0-mean GMMs for both image classes, as is usually done in natural image modeling [46] and image forensics/anti-forensics [13, 47]. Given $\mathbf{x}$, we then can form the following $2s^2M \times 1$ sized Fisher score vector:

$$\tilde{\mathbf{f}}(\mathbf{x}) = \left[\mathbf{F}(\boldsymbol{\theta}^0, \mathbf{x})^T, \mathbf{F}(\boldsymbol{\theta}^1, \mathbf{x})^T\right]^T, \tag{9}$$

which is further normalized [29, 30, 31] to construct the proposed forensic feature vector with the i-th element as:

$$\mathbf{f}_i(\mathbf{x}) = \text{sign}\left(\tilde{\mathbf{f}}_i(\mathbf{x})\right) \frac{\left|\tilde{\mathbf{f}}_i(\mathbf{x})\right|^{1/2}}{\sum_i \left|\tilde{\mathbf{f}}_i(\mathbf{x})\right|}, i = 1, 2, \cdots, 2s^2M. \tag{10}$$

In the LDR image analysis literature, the concept of Fisher scores [28] has largely influenced image classification in the form of the well-known Fisher vector [29], which further inspired the recently proposed LHS feature [30, 31] in texture/facial analysis. We use for the first time Fisher scores [28] in the proposed method for HDR image forensic purposes. We will show in Sec. 5.3 with experimental evidence that the proposed method outperforms the LHS feature. Though sharing the same basis of the Fisher vector, the proposed feature departs from the LHS feature on the following three points.

- For the proposed feature, the partial derivatives with respect to $\boldsymbol{\pi}^i$ and $(\mathbf{C}^i)^{-1}$ are calculated under *0-mean* GMMs, whereas for the LHS feature [30, 31], the partial derivatives with respect to $\boldsymbol{\mu}^i$ and $(\mathbf{C}^i)^{-1}$ are calculated using *non-0-mean* GMMs. For a fair comparison, we tried to extract the LHS feature using 0-mean GMMs. This modified LHS feature turned out to perform worse than the original LHS feature (around 4% lower detection accuracy on 8×8 image blocks). Therefore, we use the original LHS formulation for comparison with the proposed feature.

- After the partial derivative calculation and before the normalization in Eq. (10), a necessary normalization based on the mean and variance of the Fisher scores calculated on the training differential vectors is included in the construction of the LHS feature [30, 31]. However, for the proposed feature, we do not include such a procedure.

- Furthermore, differently from the LHS feature using a single GMM, the proposed method includes two GMMs for forensic analysis. The two GMMs, which respectively model mHDR images and iHDR images, provide information on both types of images. This enables the proposed feature to better distinguish between mHDR images and iHDR images than the LHS feature.

For a given HDR image preprocessed as described in Sec. 3.3, *overlapping* $s \times s$ image patches are extracted. Then, the DC component of the computed local differential vectors are removed. Afterwards, their corresponding Fisher score vectors are respectively computed according to Eq. (9). For extracting the proposed forensic feature vector of a given image, these Fisher score vectors are averaged before normalization as in Eq. (10).

Note that the Fisher score vectors are extracted and then averaged from $s \times s$ patches of a given image. Thus, for a given block size s, the feature dimensionality remains the same independently from the original image dimension. At the same time, this grants the proposed forensic feature the ability to work on image blocks as small as $s \times s$. In [48, 31], it is shown that the statistics calculated on image patches as small as 3×3 are capable of achieving good classification results. Besides, a small value of s can help to reduce the dimensionality of the feature vector. Therefore, we set $s = 3$ in this paper, and we learn one GMM from each type (mHDR, and six iHDR types: '**A**', '**B**', '**H**', '**K**', '**M**', and '**R**') from 3×3 HDR image patches. We compare the proposed method with the SPAM feature, which is 686-dimensional. Therefore, we learn GMMs with $M = 38$ components, so that the proposed forensic feature is comparably 684-dimensional. Likewise, we learn a GMM with $M = 43$ components for extracting the 688-dimensional LHS feature [30, 31]. Further in Sec. 5.3, we will also study how the number of GMM component M impacts the forensic performance of the proposed method.

5. Experimental Results

5.1. Experimental Setting

Currently, there is no large mHDR image dataset available in the literature. In order to conduct our forensic tests, we therefore collected 498 high-resolution mHDR images from the following 8 sources:

- EmpaMT dataset [49] includes 33 mHDR images;
- Meylan created 14 mHDR images [50];
- Fairchild created 106 mHDR images [51];
- HDRSID dataset [52] includes 232 mHDR images;
- IRCCyN-IVC dataset contains 10 mHDR images[1];
- Mantiuk created 8 mHDR images [53];
- Stanford dataset [54] is with 88 mHDR images;
- Ward created 7 mHDR images[2].

In order to avoid any possible intervention or postprocessing, we keep the downloaded mHDR images as they were. We rely on their authors to have adopted possible strategies [55] to create the best possible mHDR images. Our collection of downloaded mHDR images is a diverse mHDR image dataset, with images from various resources. To keep the diversity, we do not impose any mHDR image selection. In particular, we keep the mHDR images with ghosting artifacts[3] [56]. This dataset diversity can help us to better validate the robustness of forensic features.

[1] We downloaded these images from: `http://ivc.univ-nantes.fr/en/databases/ETHyma/` and `http://ivc.univ-nantes.fr/en/databases/JPEG_HDR_Images/` with duplicates removed.

[2] Included in: `http://www.cs.utah.edu/~reinhard/cdrom/hdr/`.

[3] Two example HDR images can be found here: `http://rit-mcsl.org/fairchild/HDRPS/Scenes/Peppermill.html` and `http://rit-mcsl.org/fairchild/HDRPS/Scenes/BarHarborPresunrise.html`.

Among the above mentioned 8 mHDR image sources, EmpaMT dataset and Meylan provide some original LDR images (sequences). From these two sources, we have $33+7 = 40$ well-exposed LDR images. Among them, 3 images are 24-bit RGB images in JPEG format, and the others are in RAW format. Each of these LDR images corresponds to one mHDR image. For the remaining $498-40 = 458$ mHDR images, we have no access to the original LDR image sequences. As a supplement, we randomly pick 458 images from another LDR image dataset "mitadobe5k" [57], which contains 5000 high-quality images stored in RAW format, initially used for the study of photographic global tonal adjustment. We find these LDR images especially suitable here, as their high quality ensures the creation of visually appealing iHDR images. The mHDR and LDR images are with a good diversity of scenes and cameras, which is important for reliable forensic analysis.

In our experiments, most of the LDR images in use are in the RAW format and only 3 out of 498 images are 24-bit RGB images in JPEG format. In our previous work [26], all the RAW LDR images are firstly read out to 24-bit RGB images before they are used by the inverse tone mapping algorithms to create the iHDR images. In this paper, we consider a more challenging case. We read out 16-bit pixel values from each color channel of the RAW images, then they are used to create the iHDR images. This makes the forensic detection of inverse tone mapping a much harder problem, compared to the setting in our previous work in [26]. From the results reported in this section, we can see that the performance of all the three forensic features, *i.e.*, SPAM [20], LHS [30, 31], and the proposed, decreases, if we compare with the results reported in our previous work [26]. But we will show that the new results still confirm that it is promising to perform HDR image forensics via analysis of high-order image statistics using the proposed method as well as the SPAM/LHS features.

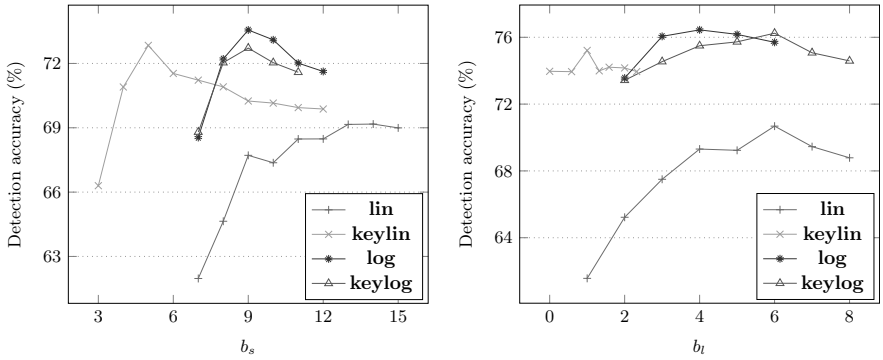
For each LDR image in our dataset, we create iHDR images '**A**', '**B**', '**H**', '**K**', '**M**', and '**R**', with the six iTM algorithms [11, 39, 33, 35, 37, 38] described in Sec. 2.2. These mHDR and iHDR images are randomly divided into a training and a test set, both containing 249 mHDR images, and 249 iHDR images of each

type ('**A**', '**B**', '**H**', '**K**', '**M**', and '**R**'). We also made sure that if iHDR images, created from one LDR image which comes from EmpaMT dataset or Meylan, are in one set, the corresponding mHDR image is also included in the same set.

Here, we choose to run forensic tests on cropped 512×512 sub-images instead of the high-resolution HDR images. This is partly because of the use of machine learning methods, *e.g.*, SVM (Support Vector Machine), which is a common practice in image forensics. In this context, a sufficient number of image samples is required to avoid the curse of dimensionality. We also need enough image samples to show the statistical significance of the tests. Though not directly testing on high-resolution images, the proposed experimental setting is very suitable to assess forensic methods. Even for high-resolution image classification under the setting of 512×512 image size, one possible way could be to simply divide the image into overlapping/non-overlapping 512×512 sub-images; the final output can then be obtained by fusing the decisions on the sub-images, *e.g.*, by majority voting.

Based on the above considerations, we cropped (at most) 9 adjacent sub-images of size 512×512 from the center of each high-resolution mHDR/iHDR image. In order to keep the number of cropped 512×512 mHDR images and that of iHDR images of each type the same, a few 512×512 iHDR images created from mitadobe5k LDR images were randomly picked and removed. In the end, we had 1851 mHDR images and 1851 iHDR images of size 512×512 in the testing dataset HDRFTE (HDR Forensic TEsting). In the training dataset HDRFTR (HDR Forensic TRaining), there are 1839 mHDR images and 1839 iHDR images sized 512×512.

Besides 512×512 images, we also choose to test on *very small* image blocks. More specifically, we consider image blocks with sizes 8×8, 7×7, 6×6, 5×5, 4×4, and even 3×3. Such tests can be taken as equivalent to fine-grained tampering localization [13], and they are very important for assessing the forensic performance of different features. From each mHDR/iHDR image, one image block is cropped randomly from each image from the HDRFTR dataset, whereas five image blocks from random locations are cropped from images in the HDRFTE



(a) SPAM [20]

(b) Proposed

Figure 2: Detection accuracy comparison for different HDR image preprocessing strategies with different parameters. The experimental study is carried out on 8×8 image blocks of mHDR and '**mix**' iHDR images. The training of the detectors is performed on HDRFTR dataset, while the accuracies are obtained from HDRFTE dataset. The legends '**lin**', '**keylin**', '**log**', and '**keylog**' respectively correspond to the HDR image scaling strategies formulated in Eqs. (2), (3), (4), and (5), by substituting b with b_s or b_l.

dataset. Therefore, for each HDR image forensic problem classifying small image blocks, we had $1839 \times 1 = 1839$ blocks from each class for training the detector, and $1851 \times 5 = 9255$ blocks from each class for testing to obtain the detection accuracies.

As described in Sec. 2.2, we have six types of iHDR images '**A**', '**B**', '**H**', '**K**', '**M**', and '**R**'. Besides, we also consider another class with mixed iHDR images randomly selected from the previously mentioned six types of iHDR images, and we denote it as '**mix**'. For each class of mHDR/iHDR images, 50 image patches with size 3×3 are randomly selected from each image in the training dataset HDRFTR. Therefore, we have $1839 \times 50 = 91950$ image patches to learn a GMM. For each image (block) size and each type of iHDR image, $1839 \times 2 = 3678$ image (block) samples from dataset HDRFTR are used for training the detector, whereas $1851 \times 2 = 3702$ image samples or $1851 \times 5 \times 2 = 18510$ image block samples from dataset HDRFTE are used to evaluate the forensic performance.

Table 1: Detection accuracy (%) comparison of different forensic detectors on 8×8 image blocks, when different HDR image preprocessing strategies are used. The b_s and b_l values are selected according to the results shown in Fig. 2. The feature dimensionalities of the SPAM feature and the proposed feature are respectively 686 and 684. Forensic detectors are trained on dataset HDRFTR. Results are obtained on dataset HDRFTE.

		mix	**A**	**B**	**H**	**K**	**M**	**R**
'**lin**'	SPAM	69.17	74.94	67.44	68.99	74.66	71.93	70.41
$b_s = 14$, $b_l = 6$	Proposed	70.68	75.34	69.88	66.86	75.69	71.10	67.37
'**keylin**'	SPAM	72.84	74.97	68.06	72.39	75.01	75.11	76.18
$b_s = 5$, $b_l = \log_2(2)$	Proposed	75.21	76.37	72.02	73.92	76.44	**76.65**	77.36
'**log**'	SPAM	73.56	74.37	71.01	73.75	74.15	74.31	79.76
$b_s = 9$, $b_l = 4$	Proposed	**76.45**	**76.70**	**75.55**	**76.21**	**76.59**	76.37	81.24
'**keylog**'	SPAM	72.72	73.39	70.24	72.77	73.46	73.71	78.60
$b_s = 9$, $b_l = 6$	Proposed	76.25	75.93	75.25	74.88	75.55	75.55	**81.90**

5.2. HDR Image Preprocessing

In this section, we will experimentally study and compare the impact of different HDR image preprocessing strategies we proposed in Sec. 3.3 on mHDR/iHDR classification accuracy. Here, we use '**lin**', '**keylin**', '**log**', and '**keylog**' to refer to the HDR image scaling strategies formulated in Eqs. (2), (3), (4), and (5), respectively. For both the SPAM feature [20] (also see Sec. 3.2) and the proposed method, we test different values of b to pre-process the HDR content, extract the forensic features, train the forensic detectors, and obtain the detection accuracies for evaluation.

We perform the experimental study on 8×8 image blocks, since the forensic performance at this level is more sensible to the quality of the employed feature. For the sake of conciseness, we report the results of this experiment for the '**mix**' iHDR class only, instead of all types of iHDR images separately, as this already includes all six types of iHDR images we consider in this paper. Fig. 2 shows the forensic performance variations with varying the b value (see Eqs. (2)-(5)) for the four proposed HDR image preprocessing strategies. In the figures, b_s

and b_l respectively denote the b values used for extracting the SPAM feature and the proposed feature.

According to Fig. 2, we obtain the best b values for different HDR image preprocessing strategies. With these values, we extract the SPAM feature and the proposed feature, and obtain the detection accuracies for other six inverse tone mapping detection problems, as reported in Table 1. In linear pixel value domain (corresponding to '**lin**' and '**keylin**'), it is interesting to observe that the scaling based on the image key performs much better than that based on the maximum value, showing that the very different dynamic ranges of different HDR images do cause bias in forensic tests. This can be improved by conducting a scaling based on the image key, which is adaptively estimated from each HDR content. Compared to the linear domain, it seems to be more effective to perform HDR image forensics in the logarithmic domain (corresponding to '**log**' and '**keylog**'). This can be interpreted by looking at the joint histogram results shown in Fig. 1. By taking the logarithm, the range of small image pixel values is expanded and thus makes histograms more distinctive. Statistical differences are well exposed in the range with small pixel values, leading to effective forensic analysis. Also, in this case normalizing by the key does not change significantly the performance, showing that a logarithmic range compression is already robust enough to cope with the content-dependent dynamic range of each image. Therefore, we choose to use in the rest of the experiments the '**log**' HDR image preprocessing strategy for both the SPAM feature and the proposed method, with $b_s = 9$ for the SPAM feature and $b_l = 4$ for the proposed feature.

5.3. Forensic Performance Evaluation

Using the '**log**' preprocessing strategy as discussed in Sec. 5.2, we report in Table 2 the forensic performance of the proposed method with comparisons to the SPAM feature [20] and the LHS feature [30, 31]. To better validate the efficacy of the proposed feature. We also carry out the forensic detection in the following two scenarios: 1) HDR images with relatively low dynamic range, and 2) HDR images with relatively high dynamic range. For each HDR image,

Table 2: Detection accuracy (%) comparison when different image (block) sizes are considered. The feature dimensionalities of the SPAM feature, the LHS feature, and the proposed feature are respectively 686, 688, and 684. Forensic detectors are trained on dataset HDRFTR. Results are obtained on dataset HDRFTE.

		'mix'	**'A'**	**'B'**	**'H'**	**'K'**	**'M'**	**'R'**
512×512	SPAM	**97.19**	**97.16**	96.92	**97.08**	**97.14**	**97.00**	**97.92**
	LHS	94.68	94.79	96.38	95.14	94.95	95.11	96.81
	Proposed	94.35	93.84	**97.11**	94.06	94.38	94.81	97.27
8×8	SPAM	73.56	74.37	71.01	73.75	74.15	74.31	79.76
	LHS	72.98	73.75	69.69	71.49	73.52	73.76	77.88
	Proposed	**76.45**	**76.70**	**75.55**	**76.21**	**76.59**	**76.37**	**81.24**
7×7	SPAM	71.39	72.25	68.07	71.26	72.37	72.36	78.36
	LHS	70.12	71.66	68.26	70.57	71.11	71.40	76.48
	Proposed	**74.67**	**74.99**	**73.95**	**74.20**	**74.79**	**74.86**	**80.32**
6×6	SPAM	69.68	70.96	66.35	70.21	70.43	70.93	75.88
	LHS	69.12	70.54	66.08	69.69	70.61	70.15	75.63
	Proposed	**72.70**	**73.33**	**71.87**	**72.01**	**73.01**	**73.12**	**78.22**
5×5	SPAM	67.10	68.25	63.24	67.26	68.17	68.50	74.23
	LHS	67.72	68.21	64.05	67.09	68.81	67.88	72.81
	Proposed	**70.82**	**71.16**	**68.64**	**70.68**	**71.38**	**71.36**	**76.52**
4×4	SPAM	63.06	63.97	59.22	63.11	63.87	64.41	69.76
	LHS	64.67	65.09	60.32	64.49	65.19	65.60	69.98
	Proposed	**67.50**	**67.71**	**64.67**	**67.25**	**67.55**	**67.91**	**73.88**
3×3	SPAM	-	-	-	-	-	-	-
	LHS	62.28	64.02	58.66	62.59	63.82	63.76	68.93
	Proposed	**63.66**	**64.89**	**60.44**	**63.92**	**64.79**	**64.63**	**70.51**

after excluding the biggest 0.01% and smallest 0.01% pixels, we can compute its dynamic range as the ratio between the biggest and smallest pixel values. According the dynamic range of each HDR image, we construct the following to subsets: 1) HDRFTE-L, containing iHDR images with the 20% lowest dynamic

Table 3: Detection accuracy (%) obtained on dataset HDRFTE-L (iHDR/mHDR images with the 20% lowest dynamic range).

		'mix'	**'A'**	**'B'**	**'H'**	**'K'**	**'M'**	**'R'**
512×512	SPAM	**98.24**	**98.11**	**98.11**	**98.11**	**97.97**	**98.51**	**98.38**
	LHS	95.54	95.95	96.62	97.30	96.08	96.89	97.57
	Proposed	95.41	93.51	96.49	94.05	93.92	95.14	96.08
8×8	SPAM	75.22	75.73	71.59	74.32	75.46	77.00	79.30
	LHS	75.92	74.95	70.70	72.30	75.24	76.32	78.76
	Proposed	**80.30**	**79.54**	**77.62**	**78.41**	**79.14**	**79.24**	**81.54**
7×7	SPAM	72.57	73.14	68.24	72.54	73.57	75.46	78.00
	LHS	73.41	73.76	70.16	73.05	73.41	74.59	77.05
	Proposed	**77.95**	**77.68**	**75.41**	**75.92**	**77.65**	**78.22**	**81.35**
6×6	SPAM	72.51	73.49	66.68	71.78	73.41	74.32	76.35
	LHS	71.97	72.84	68.24	71.67	72.62	72.78	76.57
	Proposed	**76.38**	**75.14**	**73.70**	**73.32**	**74.81**	**76.27**	**78.81**
5×5	SPAM	69.35	69.86	63.24	67.95	69.51	70.46	73.95
	LHS	69.43	69.03	65.22	68.16	69.76	70.19	73.30
	Proposed	**73.11**	**72.84**	**69.65**	**72.03**	**72.95**	**73.89**	**76.22**
4×4	SPAM	65.38	66.70	60.11	65.16	67.05	67.43	70.65
	LHS	67.84	68.16	61.16	66.59	68.49	69.46	70.65
	Proposed	**70.73**	**70.08**	**66.24**	**68.27**	**69.84**	**71.70**	**74.03**
3×3	SPAM	-	-	-	-	-	-	-
	LHS	65.30	65.32	59.86	63.76	65.38	66.57	68.41
	Proposed	**65.46**	**65.43**	**60.86**	**64.51**	**65.89**	**66.70**	**69.38**

range and mHDR images with the 20% lowest dynamic range, and 2) HDRFTE-H, containing iHDR images with the 20% highest dynamic range and mHDR images with the 20% highest dynamic range. The detection accuracies on these two subsets are reported in Tables 3 and 4 respectively.

For the extraction of both the LHS feature and the proposed feature, we use

Table 4: Detection accuracy (%) obtained on dataset HDRFTE-H (iHDR/mHDR images with the 20% highest dynamic range).

		'mix'	**'A'**	**'B'**	**'H'**	**'K'**	**'M'**	**'R'**
512×512	SPAM	**95.95**	**94.59**	93.78	**93.92**	**94.05**	**94.73**	95.68
	LHS	93.11	91.22	94.05	91.49	91.08	92.30	96.08
	Proposed	92.16	89.19	**94.32**	90.14	89.73	91.76	**97.03**
8×8	SPAM	70.54	69.38	65.05	68.97	68.43	67.73	76.73
	LHS	69.38	68.76	62.65	66.35	68.22	68.65	75.84
	Proposed	**73.51**	**69.89**	**68.73**	**69.95**	**70.46**	**69.97**	**79.46**
7×7	SPAM	68.76	68.57	61.89	67.43	68.41	68.00	77.14
	LHS	67.97	67.78	62.97	66.92	66.76	67.05	74.59
	Proposed	**72.78**	**69.70**	**68.27**	**69.00**	**70.27**	**70.38**	**78.84**
6×6	SPAM	66.86	65.59	60.30	65.57	65.43	66.41	74.54
	LHS	65.62	66.32	59.35	65.35	65.87	65.54	74.08
	Proposed	**70.16**	**69.22**	**65.22**	**67.49**	**68.35**	**69.14**	**7.30**
5×5	SPAM	64.89	64.54	57.78	63.65	64.05	64.49	73.84
	LHS	64.97	63.35	58.68	63.14	64.41	64.11	72.27
	Proposed	**68.76**	**66.65**	**61.59**	**66.54**	**66.54**	**68.65**	**76.16**
4×4	SPAM	60.43	60.41	56.19	60.59	60.05	61.11	70.32
	LHS	61.84	61.65	54.65	61.11	61.41	61.51	70.27
	Proposed	**65.70**	**64.24**	**58.27**	**63.62**	**63.59**	**64.08**	**74.08**
3×3	SPAM	-	-	-	-	-	-	-
	LHS	60.41	61.51	**56.19**	60.49	61.35	61.11	69.86
	Proposed	**62.89**	**62.92**	56.14	**62.16**	**62.35**	**62.62**	**72.92**

the same setting except the number of components of the GMMs. As described in the end of Sec. 4.2, we set the local neighborhood size $s = 3$. Furthermore, 38-component and 43-component GMMs are used respectively by the proposed method and the LHS feature, leading to a dimensionality of 684 and 688 respectively. All the forensic detectors are trained on the HDRFTR dataset using

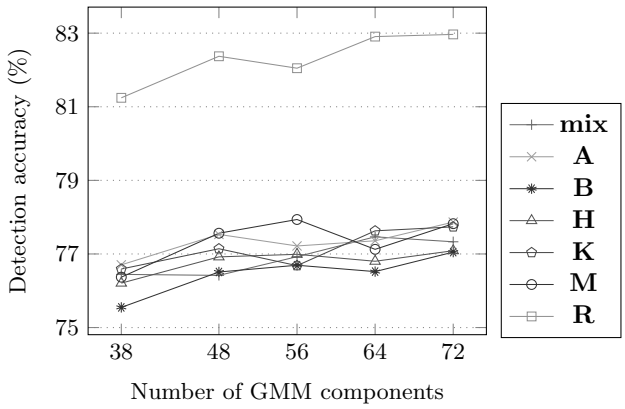


Figure 3: Detection accuracy variation of the proposed method on 8×8 image blocks, when GMMs with different numbers of components are used. Forensic detectors are trained on dataset HDRFTR. Results are obtained on dataset HDRFTE. In general, richer GMMs with more components bring higher detection accuracies.

SVM [58] with a Gaussian kernel. The parameters of the SVM are searched using a five-fold cross validation with a multiplication grid as suggested in [20].

Results reported in Tables 2-4 show the image classification on HDR images with high dynamic range is a harder problem than those with low dynamic range. In both cases, the proposed method achieves at least comparable performance with the SPAM/LHS features, and is especially advantageous on *very small* image blocks. Note that, on image blocks as small as 3×3, the SPAM feature cannot even be extracted, as it is not possible to count the co-occurrences of neighboring second-order derivatives. However, the proposed method can still perform the forensic task thanks to the fact that the GMMs are learned on 3×3 image patches. Though in such an extreme case the detection accuracies are much lower than for 512×512 images, we believe that these result show the boundary achievable by forensic methods when we keep pushing the limits of image block size. This is very important for the forensic study of *very fine-grained* image tampering localization [13].

It is possible to enrich the GMM by learning more components. One interesting question is whether richer GMMs can bring more forensic detector

power. In order to answer this question, we study the impact of the number of GMM components to the forensic performance. Besides the 38-component GMMs used to construct the 684-dimensional proposed forensic feature, we learn GMMs with 48, 56, 64, and 72 components. This leads to forensic features with higher dimensionality, *i.e.*, 864, 1008, 1152, and 1296, respectively. For different iHDR image detections, Fig. 3 shows the detection accuracy variation with respect to the number of GMM components. We can see that, in general, richer GMMs with more components can help improve the forensic performance of the proposed method. Nevertheless, 38-component GMMs seem to be already quite satisfactory for our HDR image forensic task.

6. Conclusions

This paper addresses for the first time a new forensic problem in the context of high dynamic range imaging: differentiating HDR images created from multiple LDR exposures from those obtained by single LDR images through inverse tone mapping. We point out in the paper some important and substantial differences between LDR and HDR content, *e.g.*, in terms of the physical meaning and representation of pixels. This motivates the need to study the applicability of existing forensic tools to HDR, and to eventually propose new ones. We consider as an example the popular SPAM feature and compare four HDR image preprocessing strategies to extract it from HDR images. Going one step further, we propose a more powerful HDR forensic feature, inspired by the texture/facial analysis LHS feature, by exploiting local higher-order statistics based on Fisher scores calculated under GMMs, which achieved especially competitive detection accuracies on image blocks as small as 3×3. This is especially significant when considering a scenario of fine-grained tampering localization. Experimental results show that the proposed method performs at least at a comparable level when compared to the SPAM/LHS feature, and is especially advantageous on *very small* image block classification.

As the first image forensics work in HDR content, this paper introduces

digital forensics to a new type of image, where various forensic problems can be studied. Future research shall be devoted to exploring other relevant forensic problems in HDR images, such as camera fingerprints estimation and anti-forensics. Furthermore, the proposed feature based on Fisher scores could be also successfully employed in many conventional LDR forensic problems, in order to push further the possibility of identifying image tampering on very small image blocks and gain a better understanding of the semantics of a forgery.

References

[1] P. E. Debevec, J. Malik, Recovering high dynamic range radiance maps from photographs, in: Proc. SIGGRAPH, 1997, pp. 369–378.

[2] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, K. Myszkowski, High dynamic range imaging: acquisition, display, and image-based lighting, Morgan Kaufmann, 2010.

[3] F. Banterle, A. Artusi, K. Debattista, A. Chalmers, Advanced High Dynamic Range Imaging, AK Peters / CRC Press, 2011.

[4] F. Dufaux, P. L. Callet, R. Mantiuk, M. Mrak, High Dynamic Range Video: From Acquisition, to Display and Applications, Academic Press, 2016.

[5] A. Luthra, E. François, W. Husak, Call for evidence (CfE) for HDR and WCG video coding (2015).

[6] T. Richter, On the standardization of the JPEG XT image compression, in: Picture Coding Symposium (PCS), 2013, IEEE, 2013, pp. 37–40.

[7] ITU-R, The present state of ultra-high definition television, ITU-R Recommendation BT.2246-6 (March 2017).

[8] A. Chalmers, K. Debattista, HDR video past, present and future: A perspective, Signal Processing: Image Communication 54 (2017) 49–55.

[9] F. Banterle, K. Debattista, A. Artusi, S. N. Pattanaik, K. Myszkowski, P. Ledda, A. Chalmers, High dynamic range imaging and low dynamic range expansion for generating HDR content, Comput. Graph. Forum 28 (8) (2009) 2343–2367.

[10] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, A. Vorozcovs, High dynamic range display systems, in: Proc. SIGGRAPH, 2004, pp. 760–768.

[11] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, H. H. Bülthoff, Do HDR displays support LDR content?: A psychophysical evaluation, ACM Trans. Graph. 26 (3).

[12] F. D. Simone, G. Valenzise, F. Banterle, P. Lauga, F. Dufaux, Dynamic range expansion of video sequences: a subjective quality assessment study, in: Proc. GlobalSIP, 2014, pp. 1063–1067.

[13] W. Fan, K. Wang, F. Cayre, General-purpose image forensics using patch likelihood under image statistical model, in: Proc. IEEE Int. Workshop Inf. Forensics Security, 2015, p. 6 pages.

[14] P. J. Bateman, A. T. S. Ho, J. A. Briffa, Image forensics of high dynamic range imaging, in: Proc. Digital Forensics and Watermarking, 2011, pp. 336–348.

[15] M. Stamm, K. R. Liu, Blind forensics of contrast enhancement in digital images, in: Proc. Int. Conf. on Image Processing, IEEE, 2008, pp. 3112–3115.

[16] M. C. Stamm, K. R. Liu, Forensic detection of image manipulation using statistical intrinsic fingerprints, IEEE Transactions on Information Forensics and Security 5 (3) (2010) 492–506.

[17] G. Cao, Y. Zhao, R. Ni, X. Li, Contrast enhancement-based forensics in digital images, IEEE transactions on information forensics and security 9 (3) (2014) 515–525.

[18] Z. Fan, R. L. De Queiroz, Identification of bitmap compression history: Jpeg detection and quantizer estimation, IEEE Transactions on Image Processing 12 (2) (2003) 230–235.

[19] T. Bianchi, A. Piva, Detection of nonaligned double jpeg compression based on integer periodicity maps, IEEE Transactions on Information Forensics and Security 7 (2) (2012) 842–848.

[20] T. Pevný, P. Bas, J. Fridrich, Steganalysis by subtractive pixel adjacency matrix, IEEE Trans. Inf. Forensics Security 5 (2) (2010) 215–224.

[21] J. Fridrich, J. Kodovský, Rich models for steganalysis of digital images, IEEE Trans. Inf. Forensics Security 7 (3) (2012) 868–882.

[22] T. O. Aydın, R. Mantiuk, H.-P. Seidel, Extending quality metrics to full luminance range images, in: Proc. SPIE, Vol. 6806, 2008, p. 68060B.

[23] G. Valenzise, F. De Simone, P. Lauga, F. Dufaux, A. Tescher, Performance evaluation of objective quality metrics for hdr image compression, in: Proc. SPIE, Vol. 9217, 2014, p. 92170C.

[24] A. Rana, G. Valenzise, F. Dufaux, Evaluation of feature detection in hdr based imaging under changes in illumination conditions, in: Multimedia (ISM), 2015 IEEE International Symposium on, IEEE, 2015, pp. 289–294.

[25] A. Rana, G. Valenzise, F. Dufaux, Learning-based adaptive tone mapping for keypoint detection, in: IEEE International Conference on Multimedia and Expo, IEEE, 2017, pp. 337–342.

[26] W. Fan, G. Valenzise, F. Banterle, F. Dufaux, Forensic detection of inverse tone mapping in HDR images, in: Proc. IEEE Int. Conf. Image Process., 2016, pp. 166–170.

[27] SMPTE.2084, High dynamic range electro-optical transfer function of mastering reference displays, Society of Motion Pictures and Television Engineers.

[28] T. S. Jaakkola, D. Haussler, Exploiting generative models in discriminative classifiers, in: Adv. Neural Inf. Process. Syst., 1999, pp. 487–493.

[29] F. Perronnin, J. Sánchez, T. Mensink, Improving the Fisher kernel for large-scale image classifications, in: Proc. European Conf. Comput. Vis., 2010, pp. 143–156.

[30] G. Sharma, S. ul Hussain, F. Jurie, Local higher-order statistics (LHS) for texture categorization and facial analysis, in: Proc. European Conf. Comput. Vis., 2012, pp. 1–12.

[31] G. Sharma, F. Jurie, Local higher-order statistics (LHS) describing images with statistics of local non-binarized pixel patterns, Comput. Vis. Image Und. 142 (2016) 13–22.

[32] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, H. P. A. Lensch, Optimal HDR reconstruction with linear digital cameras, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., 2010, pp. 215–222.

[33] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, G. Ward, LDR2HDR: On-the-fly reverse tone mapping of legacy video and photographs, ACM Trans. Graph. 26 (3), art. ID 39.

[34] F. Banterle, P. Ledda, K. Debattista, A. Chalmers, Inverse tone mapping, in: Proc. Int. Conf. Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, 2006, pp. 349–356.

[35] F. Banterle, P. Ledda, K. Debattista, A. Chalmers, Expanding low dynamic range videos for high dynamic range applications, in: Proc. Spring Conf. Computer Graphics, 2008, pp. 349–356.

[36] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, Photographic tone reproduction for digital images, in: Proc. SIGGRAPH, 2002, pp. 267–276.

[37] Y. Huo, F. Yang, L. Dong, V. Brost, Physiological inverse tone mapping based on retina response, Vis. Comput. 30 (5) (2014) 507–517.

[38] R. P. Kovaleski, M. M. Oliveira, High-quality reverse tone mapping for a wide range of exposures, in: Proc. SIBGRAPI, 2014, pp. 49–56.

[39] L. Meylan, S. Daly, S. Süsstrunk, Tone mapping for high dynamic range displays, in: Proc. IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging XII, 2007.

[40] IEC, 61966-2-1 multimedia systems and equipment – colour measurement and management – part 2-1: Colour management – default RGB colour space – sRGB, International Electrotechnical Commission (IEC), Geneva, Switcherland.

[41] F. Kainz, R. Bogart, Technical introduction to OpenEXR, Industrial light and magic.

[42] X. Qiu, H. Li, W. Luo, J. Huang, A universal image forensic strategy based on steganalytic model, in: Proc. ACM Int. Workshop Information Hiding and Multimedia Security, 2014, pp. 165–170.

[43] M. Kirchner, J. Fridrich, On detection of median filtering in digital images, in: Proc. IS&T/SPIE Electronic Imaging, 2010, pp. 754110:1–754110:12.

[44] T. Pouli, D. Cunningham, E. Reinhard, Statistical regularities in low and high dynamic range images, in: Proc. Symposium on Applied Perception in Graphics and Visualization, 2010, pp. 9–16.

[45] A. De Rosa, M. Fontani, M. Massai, A. Piva, M. Barni, Second-order statistics analysis to cope with contrast enhancement counter-forensics, IEEE Signal Process. Lett. 22 (8) (2015) 1132–1136.

[46] D. Zoran, Y. Weiss, Natural images, gaussian mixtures and dead leaves, in: Adv. Neural Inf. Process. Syst., 2012, pp. 1736–1744.

[47] W. Fan, K. Wang, F. Cayre, Z. Xiong, JPEG anti-forensics using non-parametric DCT quantization noise estimation and natural image statistics,

in: Proc. ACM Int. Workshop Information Hiding and Multimedia Security, 2013, pp. 117–122.

[48] M. Varma, A. Zisserman, Texture classification: Are filter banks necessary?, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., 2003.

[49] EMPA Media Technology, http://empamedia.ethz.ch/hdrdatabase/index.php.

[50] Meylan's HDR images, http://lcavwww.epfl.ch/alumni/meylan/.

[51] The HDR photographic survey, http://rit-mcsl.org/fairchild/HDR.html.

[52] High dynamic range specific image dataset (HDRSID), http://faculties.sbu.ac.ir/~moghaddam/index.php/main/page/10.

[53] HDR image gallery, http://pfstools.sourceforge.net/hdr_gallery.html.

[54] F. Xiao, J. DiCarlo, P. Catrysse, B. Wandell, High dynamic range imaging of natural scenes, in: Proc. Color Imaging Conference: Color Science, Systems, and Applications, 2002, http://white.stanford.edu/~brian/hdr/hdr.html.

[55] N. Barakat, A. N. Hone, T. E. Darcie, Minimal-bracketing sets for high-dynamic-range image capture, IEEE Trans. Image Process. 17 (10) (2008) 1864–1875.

[56] Z. G. Li, J. H. Zheng, Z. J. Zhu, S. Q. Wu, Selectively detail-enhanced fusion of differently exposed images with moving objects, IEEE Trans. Image Process. 23 (10) (2014) 4372–4382.

[57] V. Bychkovsky, S. Paris, E. Chan, F. Durand, Learning photographic global tonal adjustment with a database of input / output image pairs, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., 2011, http://groups.csail.mit.edu/graphics/fivek_dataset/.

[58] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3), Art. ID 27.